Mathematical Modeling Proposal iGEM Tec CEM Collegiate

Background

A mathematical model is a prior stage to the experimental documentation of proteins. These could include theoretical approaches in determining the protein's efficiency, the amount of proteins that will be produced and prediction of how many of these will end up being functional and capable of killing a given bacterium.

Mathematical modeling can provide essential insights to a project, and the diversity of modeling approaches and their different applications may contribute in the inference of complex experimental data.

It is crucial to count with theoretical data to put in contrast with the information obtained in experimental development; therefore the mathematical modeling have to consider several parameters in order to be a correct approximation of the expected results.

In these models three major types of approaches are applied in biological systems. These models are: thermodynamic, Boolean, and differential equation based models.

Differential equation based model

As many biological problems demand a model that can represent a multicomponent, temporally evolving dynamic system, differential equation models are arguably the best method to make an approach in order to determine the production of the Endo/Holin fused protein.

Differential equation models can be divided into two main groups: the equations using Ordinary Differential Equations (ODE), which depend on a single variable such as time, and those using Partial Differential Equations (PDE), which involve multiple variables such as time and space. For the mathematical modeling an approach using an ODE will be employed due to their advantages over PDE.

As such, differential equation models generally bring a correct approach into biological systems of moderate to high complexity, without the extreme detail of thermodynamic avenues, but with a reasonable ability to describe dynamical aspects that lack on other approaches.

One of the benefits of the Differential Equation modeling is that it is capable of representing a set of molecules such as mRNAs and proteins interacting by explicit rules defined in terms of rate equations. These equations specify the levels of each protein or

mRNA as a function of the other components as the system evolves. These models usually include time or space dependent variables such as protein and mRNA concentrations, and parameters such as production and degradation rates.

Another variable to be considered takes place when the transcription of a gene begins, in the concerning cases of proteins and RNAs, binding to regulatory sites on DNA. The frequency of this binding affects the level of expression.

Experiments have verified that a stronger binding site will increase the effect of a protein on transcription rate. On the other hand, since the DNA sequence is unchanged, the transcription is mostly determined by the amounts of transcription proteins. In translation, proteins are synthesized at ribosomes. An mRNA can be translated into one or multiple copies of corresponding proteins, which can further change the transcription of other genes.

Furthermore, transcription is mostly determined by the amounts of mRNA and translation by the amounts of proteins located in a cellular medium (intra or extra-cellular).
Feedback regulation is a term needed to explain gene expression regulation: mRNA can be translated into one or multiple copies of corresponding proteins, which can further change the transcription of other genes. [1]

Once set some conditions that differential equation models need, it is necessary to start thinking about the project so the design of the mathematical model can be created, so this model can suit the project; to have a comparative perspective, iGEM mathematical models regarding Endolysins were consulted.

The following diagram explains gene expression regulation. It is necessary to mention that this mechanism is using similar elements to the ones that will be used on the iGEM TEC CEM 2016 collegiate project.

[1] Chen, T., He, H. L. & Church, G. M. (n. d.). MODELING GENE EXPRESSION WITH DIFFERENTIAL EQUATIONS. Retrieved from http://arep.med.harvard.edu/pdf/Chen99.pdf

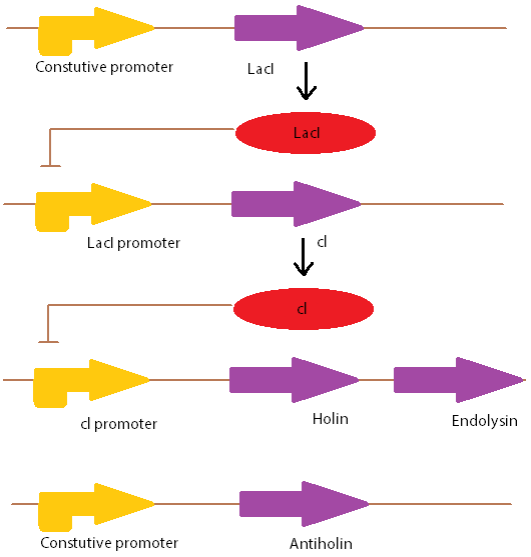

Figure 2. Gene expression regulation schematized in lacI, holin, endolysin and antiholin genes. It is worth mentioning that iGEM TEC CEM 2016 collegiate project will not be using an antiholin gene.

It can be said that endolysin-holin production is inhibited by a constitutive production of LacI, which is a protein that represses transcription of genes preceded by the cl promoter. In this case, holin and antiholin. Besides, there is a constitutive promoter that guarantees antiholin expression so that cell membrane destabilization will not happen until it is necessary. Another important thing to mention is that LacI Promoter is negatively regulated by LacI so that a transcription factor known as cl is inhibited and, thus, transcription of holin and endolysin are not allowed.

Therefore, the control mechanism of our system is inhibition of LacI by IPTG molecule. Since LacI is inhibited by IPTG (an inductor for expression of genetic fragments of interest within a vector), its effect on LacI promoter changes with respect to IPTG concentration. [2]

Holin and Endolysin expression is then controlled by cl promoter which is negatively regulated by cl. iGEM METU 2012 Team established that cell lysis (triggered by holin joined with endolysin) occurs due to higher amounts of Holin molecules (it reached 3000 molecules) rather than antiholin molecules, so that is the relevant functionality of its Kill Switch.

Now that several variables to take in consideration have been established a preliminary approach to the mathematical model and we have a comparison point for the mathematical model it is necessary to start the design of iGEM *collegiate* mathematical model.

---

[2] Team:METU/KillSwitchOverview. (n.d.). KILL-SWITCH MODEL
Retrieved from http://2012.igem.org/Team:METU/KillSwitchOverview

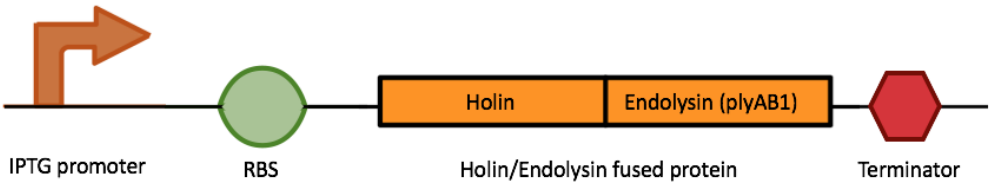

Figure 4. iGEM Tec CEM gene expression regulated by an IPTG inducible promoter.

Modeling

Objectives

- Determining The Main Equation (referred as TME), being the principal focus of TME the prediction of cell production of the Endo-holin protein.

- Determining which variables will affect TME by expressing individual equations for each of this variables, preferably in terms of time.
    - Endo-holin production and degradation rates [*Protein*]
    - mRNAs production and degradation rates [*mRNA*]

Assumptions

- Endo-holin production does not interfere with mRNA synthesis
- Ribosomes are not a limiting factor, as they are available at all times

Equations

mRNA

The rate of mRNA produced will be equal to the rate of mRNA produced minus the rate of mRNA degraded. Transcription of protein minus degradation of mRNA equals rate of change of mRNA quantity.

$$\frac{dm}{dt} = S - D_m * m$$

Where S is the transcription rate in molecules of RNA per second, D_m is the degradation rate expressed in units per seconds as it depends on how much mRNA is actually present on the cell, expressed in units of molecules of mRNA produced.

The integration of the differential equation was performed by linear integration with μ as the integrating factor.

$$\frac{dm}{dt} + D_m * m = S$$

$$\mu = e^{\int D_m dt} = e^{D_m t}$$

$$I = \int Se^{D_m t} dt = \frac{S}{D} e^{D_m t}$$

$$Y = \frac{\frac{S}{D_m} e^{D_m t} + C}{e^{D_m t}} = \frac{S}{D_m} + Ce^{-D_m t}$$

Knowing that we have an initial concentration of mRNA at time zero we stablish the condition m(0)=M_0. Substituting we obtain:

$$M_0 = \frac{S}{D_m} + C \rightarrow C = M_0 - \frac{S}{D_m}$$

We finally obtain the equation for mRNA synthesis:

$$m(t) = \frac{S}{D_m} + (M_0 - \frac{S}{D_m}) e^{-D_m t}$$

IPTG regulation in transcriptional constant

In the previous equation we considered "S" to be a constant expressed in "units of mRNA transcripted per second". However, as the gene is regulated by an IPTG induced promoter we cannot make this assumption without a valid justification.

If we want to establish an equation for mRNA production in terms of the promoter content we need to take in consideration how efficient the promoter is; this is: the nucleotide section will anneal with different strength to the transcription factors depending on which nucleotides make the actual promoter. This can be expressed in a constant called "P" for "Promoter efficiency" which will vary according to the promoter used. As we have previously stated, the promoter Bba_J04500 will be used.

The mRNA will increase in time depending on how efficient is the promoter. However, this promoter is inducible and will not be constantly active so we consider another variable "Z" as the correction factor which will determine the periods of IPTG availability. The equation will be called "n(t)", describing how much mRNA is produced in terms of the promoter efficiency.

$$n(t) = (P * t) + Z$$

Where n(t) is the amount of mRNA is produced in molecules of mRNA, P is the dimensionless efficiency of the promoter measured on a 0 to 1 scale, t is the time elapsed in minutes and Z is the dimensionless correction factor.

However, we do not need how much mRNA we produce in a determined amount of time, we need to obtain the rate of change of mRNA obtained in terms of promoter efficiency, so we differentiate.

$$\frac{dn}{dt} = P$$

We observe dn/dt equals to the promoter efficiency. This rate of change expresses how much mRNA is going to be transcripted as it is based on the promoter efficiency. Expressed

in other words: dn/dt is the transcription rate, which we have previously defined as "S". This is why we can express the variable "S" as a constant. However, said constant has not been reported yet and it should be further characterized.

Protein

The quantity of endoholin produced will be equal to the production of protein (variable T, which depends on the quantity of mRNA actually transcribed) minus the degradation of the protein (variable D_p, which depends on the quantity of protein actually present on the cell).

$$\frac{dP}{dt} = T * m - D_p * P$$

As it is shown, the protein equation relies on the mRNA equation in order to form the complete formula that will allow us to know how protein is actually formed since the transcription of mRNA. Joining the final mRNA equation into the previously exposed protein equation we obtain the following formula.

$$\frac{dP}{dt} = T * \left[\frac{S}{D_m} + (M_0 - \frac{S}{D_m}) * e^{-D_m t}\right] - D_p * P$$

Simplifying:

$$\frac{dP}{dt} = \frac{ST}{D_m} + Te^{-D_m t} * (M_0 - \frac{S}{D_m}) - D_p * P$$

$$\frac{dP}{dt} + D_p * P = \frac{ST}{D_m} + M_0 T e^{-D_m t} - \frac{STe^{-D_m t}}{D_m}$$

$$\frac{dP}{dt} + D_p * P = e^{-D_m t}\left(M_0 T - \frac{ST}{D_m}\right) + \frac{ST}{D_m}$$

Assigning simpler letters to constants:

A= D_p

B= -D_m

C=M_0T

D=ST/D_m

We obtain:

$$\frac{dP}{dt} + AP = e^{Bt}(C - D) + D$$

$$\mu = e^{\int A dt} = e^{At}$$

$$I = \int [(C - D)e^{Bt} + D]e^{At} dt = \int [(C - D)e^{Bt+At} + De^{At}]dt =$$

$$I = \int (C - D)e^{t(A+B)} dt + \int De^{At} dt$$

$$I = \frac{(C - D)}{(A + B)} e^{t(A+B)} + \frac{D}{A} e^{At}$$

$$Y = \frac{\frac{(C - D)}{(A + B)} e^{t(A+B)} + \frac{D}{A} e^{At} + K}{e^{At}} = \frac{(C - D)}{(A + B)} e^{Bt} + \frac{D}{A} + Ke^{-At}$$

Knowing that the initial condition of the protein expressed is p(0)=0 we obtain the constant:

$$0 = \frac{(C-D)}{(A+B)} + \frac{D}{A} + K \rightarrow K = \frac{-(C-D)}{(A+B)} - \frac{D}{A}$$

Our final equation results in:

$$Y = \frac{(C-D)}{(A+B)} e^{Bt} + \frac{D}{A} - \left(\frac{(C-D)}{(A+B)} + \frac{D}{A}\right) e^{-At}$$

Susbtituting the letters with the original values we obtain:

$$Y = \frac{(M_0 - \frac{ST}{D_m})}{(D_p - D_m)} e^{-D_m t} + \frac{\frac{ST}{D_m}}{D_p} - \left(\frac{(M_0 - \frac{ST}{D_m})}{(D_p - D_m)} + \frac{\frac{ST}{D_m}}{D_p}\right) e^{-D_p t}$$

Applying algebra we obtain the final expression for the construct production:

$$P(t) = \frac{M_0 T D_m - ST}{D_m(D_p - D_m)} e^{-D_m t} + \frac{ST}{D_p D_m} - \left(\frac{M_0 T D_m - ST}{D_m(D_p - D_m)} + \frac{ST}{D_p D_m}\right) e^{-D_p t}$$

Parameters

Constant translation rate

The translation rate parameter is assumed at 37°C in *E. coli.* According to the literature the translations also depend on the doublings the cell goes through while cellular division occurs and the rate of doubling. However, for practical use all parameters are assumed inside only one cell in growth times varying from 24 to 100 hours.[4]

Constant degradation rate of mRNA

The simplest model of mRNA production and the dynamics of the average level of mRNA is given by dm/dt= r - ϑm , where r is the rate of mRNA production and ϑ is the rate constant dictating mRNA decay. It is important to mention that for *E. coli*, the majority of mRNA molecules have lifetimes between 3 and 8 minutes.[3]

This time is reasonable due to the fact that every prokaryotic cell needs the needed machinery to execute its life cycle functions so that essential proteins for the microorganism are used each 20 minutes in order to accomplish physiological duties.

Constant degradation rate of protein

In *E. coli*, protein degradation plays important roles in regulating the levels of specific proteins and in eliminating damaged or abnormal proteins. In this manner protein degradation in the case that the produced Endo/Holin fused protein is not stable, this parameter depends on the composition of the protein, a low ratio leucine/methionine composition causes a variation in the half lives of proteins in *E. coli*[4].

---

[3] Bionumbers (n.d.). How fast do RNAs and proteins degrade?. Retrieved from http://www.weizmann.ac.il/plants/Milo/images/DegradationTimes120209Clean.pdf

[4] Mosteller RD, Goldstein RV, Nishimoto KR. (1979). Metabolism of individual proteins in exponentially growing Escherichia coli.Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/6987224?dopt=Abstract

Promoter efficiency

Taking into consideration a boolean system we could define our promoter as totally functional (1) or nonfunctional (0). However, our system is based on differential equations and a smoother approach is required. We take in consideration the scale 0-1 with decimals to measure our promoter's activity.

Graphics

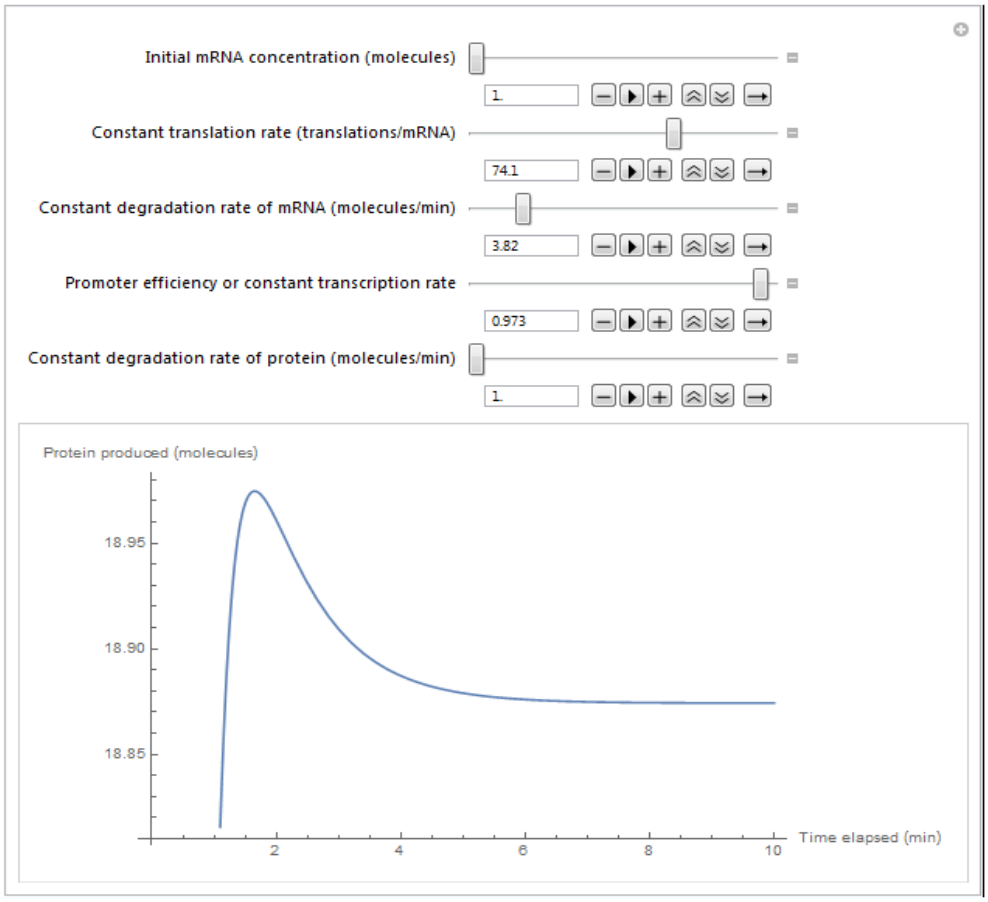

Figure 5. Mathematica model of protein expression equation with varying parameters. In this example an initial quantity of 1 mRNA molecule with 74.1 translations per mRNA is visible as well as a constant degradation rate of 3.82 molecules of mRNA degraded per minute and 1 molecule of protein degraded per minute, with a promoter almost totally efficient. This will yield a production of approximately 18 molecules of endoholin per minute once an equilibrium is reached.

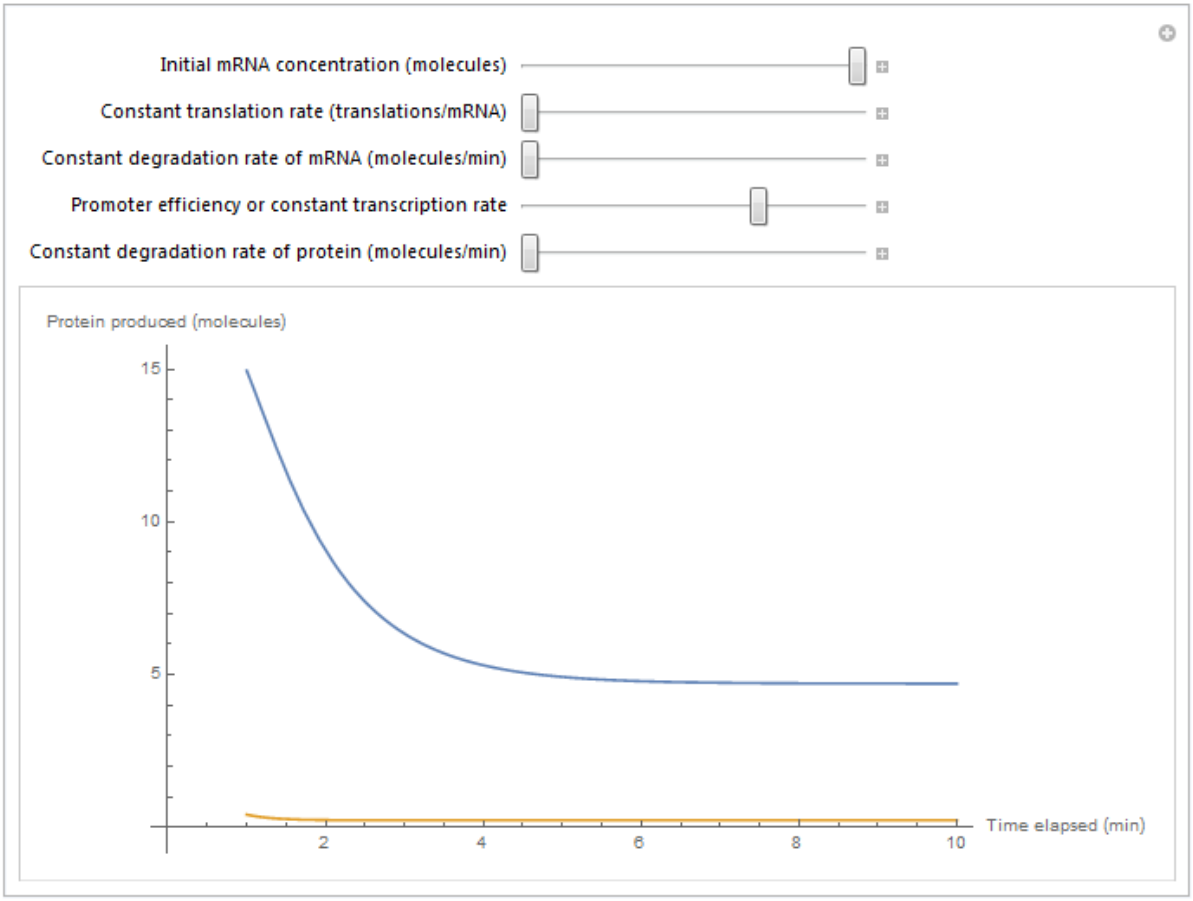

Figure 6. Mathematica model of protein expression equation and mRNA available at cell with varying parameters. In this example as we vary the parameters the mRNA (shown in orange) mantains constant at a low quantity. This is coherent with the protein production because as we increase the degradation of mRNA both lines go down.

Conclusions

By making a mathematical modeling we can predict the amount of endoholin produced, depending on several factors still to be determined. However, as this is just a proposal we consider it a viable approach because the graphics do make sense with the theoretical data, for example: if we increase the mRNA degradation rate the protein curve decreases. It is necessary to start searching for reported information to eliminate as many parameters as possible and stablish them as constants according to the dynamics of our system.